

RIGID COMPONENTS OF RANDOM GRAPHS

LOUIS THERAN

ABSTRACT. The planar rigidity problem asks, given a set of m pairwise distances among a set P of n unknown points, whether it is possible to reconstruct P , up to a finite set of possibilities (modulo rigid motions of the plane). The celebrated Maxwell-Laman Theorem from Rigidity Theory says that, generically, the rigidity problem has a combinatorial answer: the underlying combinatorial structure must contain a spanning minimally-rigid graph (Laman graph). In the case where the system is not rigid, its inclusion-wise maximal rigid substructures (rigid components) are also combinatorially characterized via the Maxwell-Laman theorem, and may be found efficiently.

Physicists have used planar combinatorial rigidity to study the phase transition between liquid and solid in network glasses. The approach has been to generate a graph via a stochastic process and then to analyze experimentally its rigidity properties. Of particular interest is the size of the largest rigid components.

In this paper, we study the emergence of rigid components in an Erdős-Rényi random graph $\mathbb{G}(n, p)$, using the parameterization $p = c/n$ for a fixed constant $c > 0$. Our first result is that for all $c > 0$, almost surely all rigid components have size 2, 3 or $\Omega(n)$. We also show that for $c > 4$, almost surely the largest rigid components have size at least $n/10$.

While the $\mathbb{G}(n, p)$ model is simpler than those appearing in the physics literature, these results are the first of this type where the distribution is over all graphs on n vertices and the expected number of edges is $O(n)$.

1. INTRODUCTION

The problem of the phase transition between liquid and solid states of glasses is an important open problem in material physics [1]. Glasses are highly disordered solids that undergo a rapid transition as they cool.

To study the phase transition, Thorpe [12] proposed a *geometric* model for the glass problem, in which bonds between the atoms are viewed as **fixed-length bars** (the bonds) connected by **universal joints** (the atoms) with full rotational degrees of freedom. Such a structure is called a **planar bar-and-joint framework** (shortly bar-joint framework, or simply framework), and these are fundamental objects of study in the field of **combinatorial rigidity** (see, e.g., [7] for a survey).

A bar-joint framework is **rigid** if the only continuous motions of the joints preserving the lengths and connectivity of the bars are rigid motions of the plane, and otherwise it is **flexible**. When a framework is flexible, it decomposes uniquely into inclusion-wise maximal rigid substructures which are called **rigid components** (shortly components); a component

2000 *Mathematics Subject Classification.* Primary 05C80; Secondary 68R10.

Key words and phrases. Computational Geometry, Rigidity Theory, Random Graphs, Rigidity Percolation, Sparse Graphs.

is non-trivial if it is larger than a single edge. In the planar case, the celebrated Maxwell-Laman Theorem [14] gives a complete characterization of *generically* minimally rigid bar-joint frameworks in terms of a combinatorial condition, which allows rigidity properties to be studied in terms of efficiently checkable graph properties.

The sequence of papers [5, 11, 12, 21, 22] studies the emergence of *large* rigid subgraphs in graphs generated by various stochastic processes, with the edge probabilities and underlying topologies used to model the temperature and chemical composition of the system. Two important observations are that: (1) very large rigid substructures emerge very rapidly; (2) the transition appears to occur slightly below average degree 4 in the the planar bar-joint model.

Main result novelty. In this paper, we study the emergence of rigid components in random graphs generated by a simple, well-known stochastic process: the Erdős-Rényi random graph model $\mathbb{G}(n, p)$, in which each edge is included with probability p , independently. We consider edge probabilities of the form $p = c/n$, where c is a fixed constant, and consider the size of the largest rigid components in $\mathbb{G}(n, p)$.

Our main result is the following statement about rigid components in $\mathbb{G}(n, c/n)$.

Theorem 1 (Size and emergence of a large rigid component). *Let $c > 0$ be a constant. Almost surely, all rigid components in $\mathbb{G}(n, c/n)$ span 2, 3, or $\Omega(n)$ vertices. If $c > 4$, then almost surely there are components of size at least $n/10$.*

(A random graph has a property almost surely if the probability of $\mathbb{G}(n, p)$ having it tends to one as $n \rightarrow \infty$.)

To the best of our knowledge, this is the first proven result on the emergence of rigid components in random graphs that have, almost surely, close to $2n - 3$ edges (the number required for minimal rigidity) but *no other special assumptions*, such as being highly connected or a subgraph of a hexagonal lattice, both of which play critical roles in the previous results on the rigidity of random graphs.

It is important to note that rigidity is inherently a non-local phenomenon: adding a single edge to a graph that has no non-trivial rigid components may rigidify the entire graph (or removing a single edge may cause a large rigid component to shatter). It is this property of rigidity that distinguishes it from the well-studied k -core problem in random graph theory.

In the proof of Theorem 1, we formalize the experimental observation that rigid components, once they appear, are very likely to grow rapidly. Although the proof of Theorem 1 relies mainly on standard tools for bounding sums of independent random variables, our result seems to be the first that directly analyzes rigidity properties of $\mathbb{G}(n, p)$, rather than reducing to a connectivity property.

Related work. Jackson, et al. [10] studied the space of random 4-regular graphs and showed that they are almost surely globally rigid (see [6, 9]). They also established a threshold for $\mathbb{G}(n, p)$ to be rigid at $p = n^{-1}(\log n + 2 \log \log n + \omega(1))$, which coincides with the threshold for $\mathbb{G}(n, p)$ to almost surely have all vertices with degree at least 2. The approach in [10] is based on combining results on the connectivity of random graphs (e.g., [17, Theorem 4]) and theorems linking rigidity and connectivity proved in [10] and also [9, 16]. In the $\mathbb{G}(n, p)$ model, the techniques there seem to rely on the existence of a very large 6-core, so it does not seem that they can be easily adapted to our setting when c is close to 4 (below the threshold for even the 4-core to emerge [19]).

Holroyd [8] extended the formal study of connectivity percolation [3] to rigidity percolation in the hexagonal lattice. He shows, via a reduction to connectivity percolation, that there is an edge-probability threshold for the existence of an infinite¹ rigid component in the hexagonal lattice which is higher than that for connectivity. It is also shown in [8] that the infinite component, when it exists, is unique for all but a countable set of edge probabilities p . All the proofs in [8] rely in an essential way on the structure of the hexagonal lattice (in particular that a suitably defined tree in its dual graph is a dual of a rigid component).

The fundamental k -core problem in random graph theory has been studied extensively, with a number of complete solutions. Łuczak [17] first proved that for $k \geq 3$, the (it is always unique, if present) k -core is, almost surely, either empty or has linear size. Pittel, et al. solved the k -core problem, giving an exact threshold for its emergence and bounds on its size [19]. Janson and Łuczak gave an alternative proof of this result, using simpler stochastic processes [13]. All these results are based on analyzing a process that removes low-degree vertices one at a time, which does not apply in the rigidity setting.

2. PRELIMINARIES

In this section we give the technical preliminaries required for the proof of Theorem 1.

Combinatorial rigidity. An **abstract bar-and-joint framework** (G, ℓ) is a graph $G = (V, E)$ and vector of non-negative **edge lengths** $\ell = \ell_{ij}$, for each edge $ij \in E$. A realization $G(\mathbf{p})$ of the abstract framework (G, ℓ) is an embedding of G onto the planar point set $\mathbf{p} = (\mathbf{p}_i)_1^n$ with the property that for all edges $ij \in E$, $\|\mathbf{p}_i - \mathbf{p}_j\| = \ell_{ij}$. The framework (G, ℓ) is **rigid** if it has only a discrete set of realizations modulo trivial plane motions, and is **flexible** otherwise.

A graph $G = (V, E)$ is **(2, 3)-sparse** if every subgraph induced by $n' \geq 2$ vertices has at most $2n' - 3$ edges. If, in addition, G has $2n - 3$ edges, G is **(2, 3)-tight** (shortly, Laman).

The Maxwell-Laman Theorem completely characterizes the rigidity of generic planar bar-joint frameworks.

Proposition 2 (Maxwell-Laman Theorem [14]). *A generic bar-joint framework in the plane is minimally rigid if and only if its graph is (2, 3)-tight.*

Genericity is a subtle concept, and we refer the reader to our paper [20] for a detailed discussion. In the following it suffices to note that for a fixed G almost all \mathbf{p} are generic, and that, by the Maxwell-Laman Theorem, all generic frameworks $G(\mathbf{p})$ have the same rigidity properties.

If G contains a spanning Laman graph it is **(2, 3)-spanning** (shortly rigid). A rigid induced subgraph is called a **spanning block** (shortly block), and an inclusion-wise maximal block is a **spanning component** (shortly component)². By [15, Theorem 5], every graph decomposes uniquely into components, and every edge is spanned by exactly one component. A component is non-trivial if it contains more than one edge. Figure 1(a) shows an example of a Laman graph. Figure 1(b) has an example of a flexible graph with its components indicated; they are the two triangles and two trivial components consisting of a single edge only.

¹Rigidity of infinite frameworks is a subtle concept, and [8] devotes careful attention to its development.

²In [15] the terms “block” and “component” are reserved for induced subgraphs of Laman graphs, but there is no concern of confusion here.

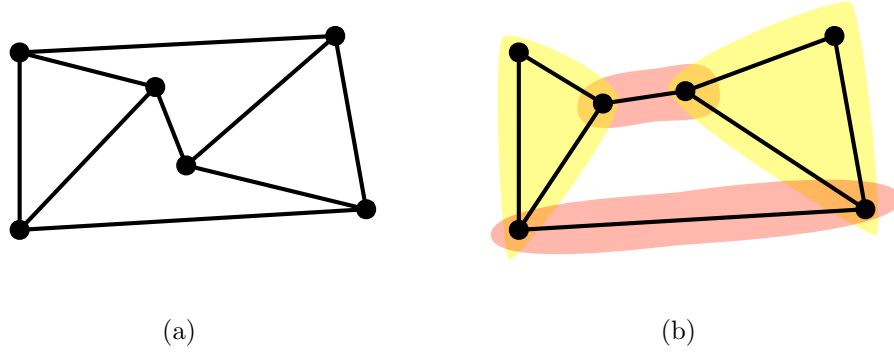


FIGURE 1. Laman graphs and rigid components: (a) a Laman graph on $n = 6$ vertices; (b) a flexible graph with its rigid components indicated.

An alternative characterization of Laman graphs is via so-called **Henneberg constructions**, which are local moves that transform Laman graphs on n vertices to Laman graphs on $n + 1$ vertices (see [15, Section 6]). The **Henneberg I** move adds a new vertex n to a Laman graph G and attaches it to two neighbors in $V(G) - n$. It is a fundamental result of rigidity theory that the Henneberg I move preserves generic rigidity [14]³.

We summarize the properties of rigid graphs and components that we will use below in the following proposition.

Proposition 3 (Properties of rigid graphs and rigid components). *Let $G = (V, E)$ be a simple graph with n vertices.*

- (a) *G decomposes uniquely into rigid components (inclusion-wise maximal induced Laman graphs), and every edge is in some component [15, Theorem 5].*
- (b) *Adding an edge to a graph G never decreases the size of any rigid component [15, Theorem 2].*
- (c) *If G' is a block in G with vertices $V' \subset V$ and there is a vertex $i \notin V'$ with at least two neighbors in V' , then G' is not a component of G .*
- (d) *If G has at least $2n - 2$ edges, then it contains a component spanning at least 4 vertices [15, Theorem 2 and Theorem 5].*

What we have presented here is a small part of a well-developed combinatorial and algorithmic theory of (k, ℓ) -sparse graphs. We refer the reader to [15] for a detailed treatment of the rich properties of sparse graphs.

Tools from random graph theory. One of our main technical tools is the following result on the size of dense subgraphs in $\mathbb{G}(n, c/n)$ due to Łuczak [17]. Since it appears without proof in [17], we give our own in the appendix.

Proposition 4 (Density Lemma [17]). *Let a and c be real constants with $a > 1$ and $c > a$. Almost surely, $\mathbb{G}(n, c/n)$ has no subgraphs with at most $k = t(a, c)n$ vertices and at least akn edges, where*

$$t(a, c) = \left(\frac{2a}{c} \right)^{\frac{a}{a-1}} e^{-\frac{a+1}{a-1}}$$

³This fact, along with an analogous result for the so-called Henneberg II move, which adds a vertex of degree 3, is the core of Laman's proof.

We will also make use of a fairly general form of the Chernoff bound for the upper tail of the binomial.

Proposition 5 (Chernoff bound). *Let $\text{Bin}(N, p)$ be a binomial random variable with parameters n and p . Then for all $\delta > 0$,*

$$\Pr[\text{Bin}(N, p) \geq (1 + \delta)Np] \leq \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{Np}$$

Large deviation bounds of this type are attributed to Chernoff [4], and are standard in combinatorics. The specific form of Proposition 5 appears in, e.g., [18, Theorem 4.1, p. 68].

3. PROOFS

In this section we prove the main result of this paper.

Theorem 1 (Size and emergence of a large rigid component). *Let $c > 0$ be a constant. Almost surely, all rigid components in $\mathbb{G}(n, c/n)$ span 2, 3, or $\Omega(n)$ vertices. If $c > 4$, then almost surely there are components of size at least $n/10$.*

Proof outline. Here is the proof strategy in a nutshell. Because any rigid component with $n' \geq 4$ vertices must be somewhat dense, the very general bound of Proposition 4 implies that for $p = c/n$ all the components are either trivial, triangles, or spanning a constant fraction of the vertices in $\mathbb{G}(n, c/n)$ (Lemma 6). We then improve upon our bounds on the probability of components of size sn , for $s \in (0, 1)$ by formalizing the observation that such components are likely to “grow” (Lemma 8) and then optimizing s (Lemma 9).

The rest of this section contains the details.

Rigid components have either constant or linear size. We start by proving that non-trivial rigid components are all very large or triangles, almost surely.

Lemma 6. *Let $c > 0$ be a fixed constant. Almost surely, all rigid components in $\mathbb{G}(n, c/n)$ have size 2, 3, or $\Omega(n)$.*

Proof. By Proposition 3(a), any rigid component on $n' \geq 4$ vertices has at least $\frac{5}{4}n'$ edges (with equality for $n' = 4$). The lemma then follows from Proposition 4 and the well-known fact that almost surely $\mathbb{G}(n, c/n)$ contains a triangle [2, Theorem 4.1, p. 79]. \square

Remark: In fact, this proof via Proposition 4 implies a stronger result, which is that almost surely $\mathbb{G}(n, c/n)$ does not contain any sub-linear size induced subgraphs with enough edges to be non-trivial rigid blocks, except for triangles.

For $c > 4$, the number of edges in $\mathbb{G}(n, c/n)$ implies that it has at least one large rigid component, almost surely.

Lemma 7. *Let $c > 4$. Almost surely, $\mathbb{G}(n, c/n)$ contains at least one component of size $\Omega(n)$.*

Proof. For any $\epsilon > 0$ $\mathbb{G}(n, (4 + \epsilon)/n)$ has at least $2n - 2$ edges with high probability. Proposition 3(d) then implies that almost surely $\mathbb{G}(n, (4 + \epsilon)/n)$ contains at least one rigid component with at least 4 vertices. By Lemma 6, all of these span at least $t(a, 4 + \epsilon)n$ vertices.

By Proposition 3(b) the size of rigid components is an increasing property and [2, Theorem 2.1, p. 36], this lower bound on size holds, almost surely, for any $c > 4$. \square

For $c > 4$ the largest component is very large. We now turn to improving the lower bound on the size of rigid components. To do this, we will use the maximality of components as well as their edge density.

Lemma 8. *The probability that a fixed set of k vertices spans a component in $\mathbb{G}(n, c/n)$ is at most*

$$(1) \quad \Pr [\text{Bin}(k^2/2, c/n) \geq 2k - 3] \left((1 - c/n)^k + k \frac{c}{n} (1 - c/n)^{k-1} \right)^{n-k}$$

Proof. To induce a component, a set V' of k vertices must span at least $2k - 3$ edges by Proposition 3(a). By Proposition 3(c) if V' spans a component, no vertex outside of V' can have more than one neighbor in V' . The two terms in (1) correspond to these two events, which are independent. \square

Remark: This estimate of the probability of a set of vertices inducing a component is very weak, since it uses only the number of edges induced by V' (not their distribution) and the simplest local obstacle to maximality. Any improvement in this part of the argument would translate into improvements in the lower bound on the size of components.

Lemma 9. *For $c > 4$, almost surely all components span at least $n/10$ vertices.*

Proof. With the assumptions of the lemma, by Lemma 7, $\mathbb{G}(n, c/n)$ almost surely has no blocks of size smaller than tn , where t is a constant independent of n . It follows from Proposition 3(a) that $\mathbb{G}(n, c/n)$ almost surely has no components smaller than tn .

Let X_k to be the number of components of size k and let s be a parameter to be selected later. We will show that $\sum_{k=4}^{sn} \mathbf{E}[X_k] = o(1)$, which implies the lemma by a Markov's inequality. As noted above, $\sum_{k=4}^{tn} \mathbf{E}[X_k] = o(1)$, so we concentrate on $k \in [tn, sn]$.

By Lemma 8

$$\begin{aligned} \mathbf{E}[X_k] &\leq \binom{n}{k} \Pr [\text{Bin}(k^2/2, c/n) \geq 2k - 3] \left((1 - c/n)^k + k \frac{c}{n} (1 - c/n)^{k-1} \right)^{n-k} \\ &\leq \left(\frac{en}{k} \right)^k \Pr [\text{Bin}(k^2/2, c/n) \geq 2k] \left((1 - c/n)^k + k \frac{c}{n} (1 - c/n)^{k-1} \right)^{n-k} + o(1) \end{aligned}$$

Setting $k = sn$ and letting $c = 4 + \epsilon$, we use the Chernoff bound to obtain

$$\begin{aligned} &\left(\frac{e}{s} \right)^{sn} \Pr [\text{Bin}(k^2/2, (4 + \epsilon)/n) \geq 2sn] \left((1 - (4 + \epsilon)/n)^s n + cs(1 - (4 + \epsilon)/n)^{sn-1} \right)^{n-sn} \leq \\ &\left(\frac{e}{s} \right)^{sn} \left(e^{\frac{-\epsilon s - 4s + 4}{s(\epsilon + 4)}} \left(\frac{-\epsilon s - 4s + 4}{s(\epsilon + 4)} + 1 \right)^{-\frac{-\epsilon s - 4s + 4}{s(\epsilon + 4)} - 1} \right)^{\frac{1}{2} ns^2 (\epsilon + 4)} (e^{-(4 + \epsilon)s} (1 + (4 + \epsilon)s))^n \end{aligned}$$

As $\epsilon \rightarrow 0$ the right-hand side approaches

$$e^{ns} \left(e^{\frac{1}{s} - 1} \left(\frac{1}{s} \right)^{-1/s} \right)^{2ns^2} \left(\frac{1}{s} \right)^{ns} (e^{-4s} (4s + 1))^{n - ns}$$

Substituting $s = 1/10$, this simplifies to

$$2^{-n/10} 5^{-n} 7^{9n/10} e^{-2n/25} = e^{-\Theta(n)}$$

(which can be seen by taking the logarithm and factoring out n). Since this bound is good for any $s' \in [t, 1/10]$, we have $\sum_{k=tn}^{n/10} \mathbf{E}[X_k] \leq n e^{-\Theta(n)} = o(1)$.

By Proposition 3(b) the size of rigid components is an increasing property and [2, Theorem 2.1, p. 36], this lower bound on size holds almost surely for any $c > 4$. \square

4. CONCLUSIONS AND OPEN PROBLEMS

We considered the question of the size and emergence of rigid components in a random graph $\mathbb{G}(n, c/n)$ as c increases, and we proved that almost surely all rigid components in $\mathbb{G}(n, c/n)$ are single edges, triangles or span $\Omega(n)$ vertices. For $c > 4$, we proved that, almost surely, the largest rigid components span at least $n/10$ vertices.

The most natural open question is whether there is a threshold constant for rigid components in $\mathbb{G}(n, p)$.

Question 10 (Existence of a threshold constant). *Is there a constant c_r at which a linear-sized rigid component appears in $\mathbb{G}(n, (c_r + \epsilon)/n)$ almost surely, and $\mathbb{G}(n, (c_r - \epsilon)/n)$ almost surely has no large rigid components?*

The other important question is about the structure of large rigid components when they emerge.

Question 11 (Structure of large rigid components in $\mathbb{G}(n, c/n)$). *Is there almost surely only one large rigid component in $\mathbb{G}(n, c/n)$, and what are the precise bounds on its size?*

We have observed in computer simulations that when linear sized rigid components are present, there is only one, and it is much larger than $n/10$.

REFERENCES

- [1] P. Anderson. Through the glass lightly. *Science*, 267(5204):1615–1616, 1995.
- [2] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001. ISBN 0-521-80920-7; 0-521-79722-5.
- [3] B. Bollobás and O. Riordan. *Percolation*. Cambridge University Press, New York, 2006. ISBN 978-0-521-87232-4; 0-521-87232-4.
- [4] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952. ISSN 0003-4851.
- [5] M. V. Chubynsky and M. F. Thorpe. Rigidity percolation and the chemical threshold in network glasses. *J. Optoelectron. Adv. Mater.*, pages 229–240, 2002. Festschrift honoring Stanford Robert Ovshinsky on his eightieth birthday. Includes bibliographical references and indexes.
- [6] R. Connelly. Generic global rigidity. *Discrete and Computational Geometry*, 33(4): 549–563, April 2005. doi: DOI:10.1007/s00454-004-1124-4.
- [7] J. Graver, B. Servatius, and H. Servatius. *Combinatorial rigidity*. Graduate Studies in Mathematics. American Mathematical Society, 1993.
- [8] A. E. Holroyd. Existence and uniqueness of infinite components in generic rigidity percolation. *Annals of Applied Probability*, 8(3):944–973, 1998.
- [9] B. Jackson and T. Jordán. Connected rigidity matroids and unique realizations of graphs. *Journal of Combinatorial Theory Series B*, 94(1):1–29, May 2005.
- [10] B. Jackson, B. Servatius, and H. Servatius. The 2-dimensional rigidity of certain families of graphs. *Journal of Graph Theory*, 54(2):154–166, 2007. ISSN 0364-9024.

- [11] D. J. Jacobs and B. Hendrickson. An algorithm for two-dimensional rigidity percolation: the pebble game. *Journal of Computational Physics*, 137:346–365, November 1997.
- [12] D. J. Jacobs and M. F. Thorpe. Generic rigidity percolation: The pebble game. *Phys. Rev. Lett.*, (75):4051–4054, 1995.
- [13] S. Janson and M. J. Luczak. A simple solution to the k -core problem. *Random Structures & Algorithms*, 30(1-2):50–62, 2007. ISSN 1042-9832.
- [14] G. Laman. On graphs and rigidity of plane skeletal structures. *Journal of Engineering Mathematics*, 4:331–340, 1970.
- [15] A. Lee and I. Streinu. Pebble game algorithms and sparse graphs. *Discrete Mathematics*, 308(8):1425–1437, April 2008. <http://dx.doi.org/10.1016/j.disc.2007.07.104>.
- [16] L. Lovász and Y. Yemini. On generic rigidity in the plane. *SIAM J. Algebraic and Discrete Methods*, 3(1):91–98, 1982.
- [17] T. Luczak. Size and connectivity of the k -core of a random graph. *Discrete Math.*, 91(1):61–68, 1991. ISSN 0012-365X.
- [18] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995. ISBN 0-521-47465-5.
- [19] B. Pittel, J. Spencer, and N. C. Wormald. Sudden emergence of a giant k -core in a random graph. *Journal of Combinatorial Theory. Series B*, 67(1):111–151, 1996. ISSN 0095-8956.
- [20] I. Streinu and L. Theran. Combinatorial genericity and minimal rigidity. In *SCG '08: Proceedings of the twenty-fourth annual symposium on Computational Geometry*, pages 365–374, New York, NY, USA, 2008. ACM.
- [21] M. F. Thorpe and M. V. Chubynsky. Self-organization and rigidity in network glasses. *Current Opinion in Solid State & Materials Science*, 5:525–532, 2002.
- [22] M. F. Thorpe, D. J. Jacobs, M. V. Chubynsky, and A. J. Rader. Generic rigidity of network glasses. In *Rigidity Theory and Applications*, pages 239–277. Kluwer Academic/Plenum Publishing, NY., 1999.

APPENDIX A. PROOF OF PROPOSITION 4

We give a proof of Proposition 4, which specializes a technical lemma from [17]. It appeared there without proof.

Proposition 11 (Density Lemma [17]). *Let a and c be real constants with $a > 1$ and $c > a$. Almost surely, $\mathbb{G}(n, c/n)$ has no subgraphs with at most $k = t(a, c)n$ vertices and at least akn edges, where*

$$t(a, c) = \left(\frac{2a}{c} \right)^{\frac{a}{a-1}} e^{-\frac{a+1}{a-1}}$$

Proof. Let $t = t(n)$ be a parameter to be picked later, and say that a subgraph is bad if it has at most tn vertices and at least atn edges, and let X be the number of bad subgraphs in $\mathbb{G}(n, c/n)$. The proof is via a first moment argument. We observe that for any set of k vertices in $\mathbb{G}(n, c/n)$, the number of induced edges is a random variable that is dominated by the binomial random variable $\text{Bin}(k^2/2, c/n)$.

Let X_k be the number of bad subgraphs of size k . By definition,

$$(2) \quad \mathbf{E}[X] \leq \sum_{k=2}^{tn} \mathbf{E}[X_k] \leq \sum_{k=2}^{tn} \binom{n}{k} \sum_{j=ak}^{k^2/2} \binom{k}{j} \left(\frac{c}{n} \right)^j \left(1 - \frac{c}{n} \right)^{k-j}$$

and we will show that the right hand side $o(1)$, for a choice of t independent of n , which implies the lemma since Markov's inequality shows that $\mathbf{Pr}[X > 0] \leq \mathbf{E}[X]$. To do this, we split the sum in (2) into two parts: $2 \leq k \leq n^\epsilon$, where $\epsilon < \min\{\frac{2a(1-1/a)}{a+3}, 1/2\}$; and $n^\epsilon < k \leq tn$.

For the small terms, we start by expanding $\mathbf{E}[X_k]$ directly:

$$\begin{aligned}
\sum_{k=2}^{n^\epsilon} \binom{n}{k} \sum_{j=ak}^{k^2/2} \binom{k^2/2}{j} \left(\frac{c}{n}\right)^j \left(1 - \frac{c}{n}\right)^{k-j} &\leq \sum_{k=2}^{n^\epsilon} \left(\frac{en}{k}\right)^k \sum_{j=ak}^{k^2/2} \left(\frac{ek}{2j}\right)^j \left(\frac{c}{n}\right)^j \\
&\leq \sum_{k=2}^{n^\epsilon} \left(\frac{en}{k}\right)^k \sum_{j=ak}^{k^2/2} \left(\frac{ekc}{2an}\right)^{ak} \\
&\leq \sum_{k=2}^{n^\epsilon} \sum_{j=ak}^{k^2/2} \left(\frac{k^{ak}}{k^k}\right) \left(\frac{e^{1+1/a}c}{2an^{1-1/a}}\right)^{2a} \\
&\leq n^{\epsilon(3+a)} \left(\frac{e^{1+1/a}c}{an^{1-1/a}}\right)^{2a} \\
&= o(1)
\end{aligned}$$

For $k > n^\epsilon$ we parameterize k as tn with $t > n^{\epsilon-1}$ and use the Chernoff inequality to bound the probability that $\text{Bin}(\frac{1}{2}(tn)^2, c/n) > atn$. Plugging in to Proposition 5 with $\delta = \frac{2a}{ct} - 1$ shows that the probability of any particular set of tn vertices inducing a bad subgraph is at most

$$(3) \quad \left(e^{\frac{2a}{ct}-1} \left(\frac{2a}{ct} \right)^{-\frac{2a}{ct}} \right)^{\frac{1}{2}cnt^2}$$

The number of sets of size tn is at most $(e/t)^{tn}$. Multiplying with (3) gives a bound on $\mathbf{E}[X_{tn}]$:

$$\mathbf{E}[X_{tn}] \leq \left(e^{(a+1)t - \frac{ct^2}{2}} \left(\frac{1}{t} \right)^t \left(\frac{2a}{ct} \right)^{-at} \right)^n$$

We can show that t can be chosen independently of n to make the inner expression strictly less than one. Taking the logarithm, we obtain

$$t \left(a + 1 - \frac{ct}{2} - a \log \left(\frac{2a}{c} \right) + (a-1) \log(t) \right)$$

Plugging in $t(a, c)$ from the statement, this simplifies to $-\frac{ct(a, c)^2}{2} < 0$, from our assumptions on a and c , and this function decreases with t in the interval $[n^{\epsilon-1}, t(a, c)]$.

It follows that

$$\sum_{k=n^\epsilon}^{t(a, c)n} \mathbf{E}[X_k] \leq ne^{-\Theta(n)} = o(1)$$

completing the proof. □

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF MASSACHUSETTS, AMHERST

E-mail address: `theran@cs.umass.edu`

URL: `http://cs.umass.edu/~theran/`